

# Design superiority of palindromic DNA sites for site-specific recognition of proteins: Tests using protein stitchery

CHANGMOON PARK\*†, JUDY L. CAMPBELL†‡, AND WILLIAM A. GODDARD III\*†§

\*Materials and Molecular Simulation Center, Beckman Institute (139-74), †Division of Chemistry and Chemical Engineering, and ‡Division of Biology, California Institute of Technology, Pasadena, CA 91125

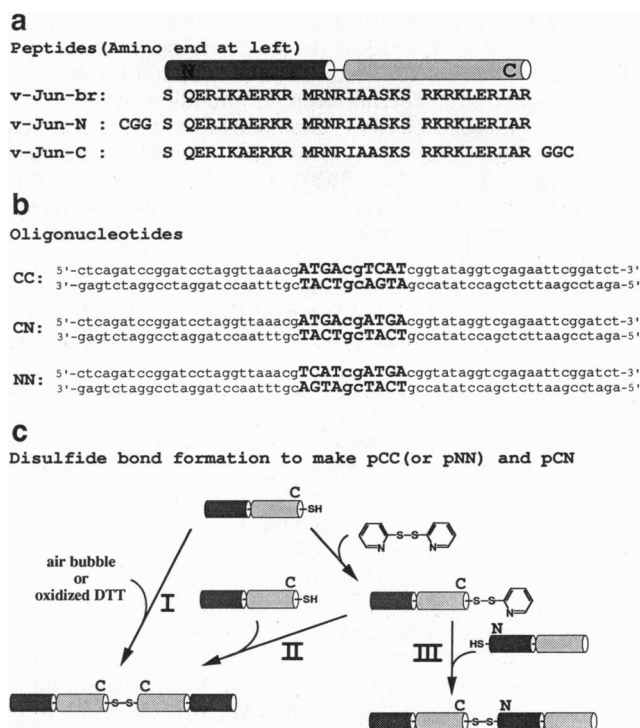
Contributed by William A. Goddard III, February 12, 1993

**ABSTRACT** Using protein stitchery with appropriate attachment of cysteines linking to either C or N termini of the basic region of the v-Jun leucine zipper gene-regulatory protein, we constructed three dimers—pCC, pCN, and pNN. All three bind specifically to the appropriately rearranged DNA recognition sites for v-Jun: ATGAcgTCAT, ATGAcgATGA, and TCATcgTCAT, respectively ( $K_d$ ,  $\approx 4$  nM at 4°C). Results of DNase I footprinting provide strong support for bent recognition helices in leucine zipper protein–DNA complexes. Comparison of the results for pCC and pNN with those for pCN shows the design superiority of palindromic sequences for protein recognition.

The mechanism by which cells respond to external stimuli is a fundamental problem in modern biology. Transcriptional regulatory proteins are known to play a key role in several systems evolved by cells to convert extracellular signals into altered gene expression (1). They operate by specifically binding to DNA target sites, which regulate the transcription of particular genes. Prominent among transcriptional regulatory proteins are the leucine zipper family of proteins, which recognize the DNA binding site as either homodimers or heterodimers (2–4).

The leucine zipper proteins are characterized by two functional segments: (i) the leucine zipper region, a helical region containing four or five leucines spaced at seven-amino acid intervals, and (ii) the basic region containing many basic residues (5–10). The basic region appears to be unfolded in solution but assumes an  $\alpha$ -helical structure binding to its recognition site (11–13). Site-directed mutagenesis (6, 7) and domain swapping (8–10) experiments show that the leucine zipper region mediates dimerization and that the basic region is responsible for DNA binding. Experiments replacing the leucine zipper region with a three-peptide linker (14, 15) showed that the dimerized basic region recognizes the same site as the native protein, supporting the scissors grip model (5), where each monomer recognizes the half site of the symmetrical DNA binding site. Recently, we showed that the normal dimer (denoted pCC), which selectively recognizes the sequence ATGAcgTCAT, can be inverted to form a protein (denoted pNN) that selectively recognizes the inverted site, TCATcgATGA (15).

Gel electrophoresis experiments (22) with Jun homodimer and with Jun–Fos heterodimer showed that Jun and Fos induce DNA bending in the opposite direction upon binding to the specific site. To explain this, it was proposed that the basic region of Jun has a bent  $\alpha$ -helix, while the basic region of Fos has a straight helix. However, a recent x-ray crystal structure (21) for the complex between GCN4 and DNA containing the GRE site (ATGACTCAT) showed a straight single  $\alpha$ -helix for the basic region of GCN4. Our current



**FIG. 1.** Sequences of protein (a) and oligonucleotides (b) used in gel-retardation and footprinting studies. Total length of each oligonucleotide is 62. v-Jun-br contains the basic region of v-Jun. CGG is added to the N terminus of v-Jun-br to make v-Jun-N and GGC is added to the C terminus of v-Jun-br to make v-Jun-C. Proteins were chemically synthesized and checked by mass spectroscopy at the Biopolymer Synthesis Center at the California Institute of Technology (15). (c) Strategy for making pCC (and pNN) and pCN dimers. v-Jun-C was incubated with 10 mM dithiothreitol (DTT) for 5 hr at room temperature and purified directly into 10 mM 2,2'-dithiodipyridine/100 mM sodium phosphate, pH 5.5, containing 30% acetonitrile. Resulting thiopyridyl(v-Jun-C) was purified by HPLC. Purified monomer v-Jun-N underwent reaction with 2 equivalents of thiopyridyl(v-Jun-C) in solution containing 100 mM tetraethylammonium acetate buffer (pH 7.5) and 15% acetonitrile for 12 hr at room temperature. The final product, pCN, was purified by HPLC (15).

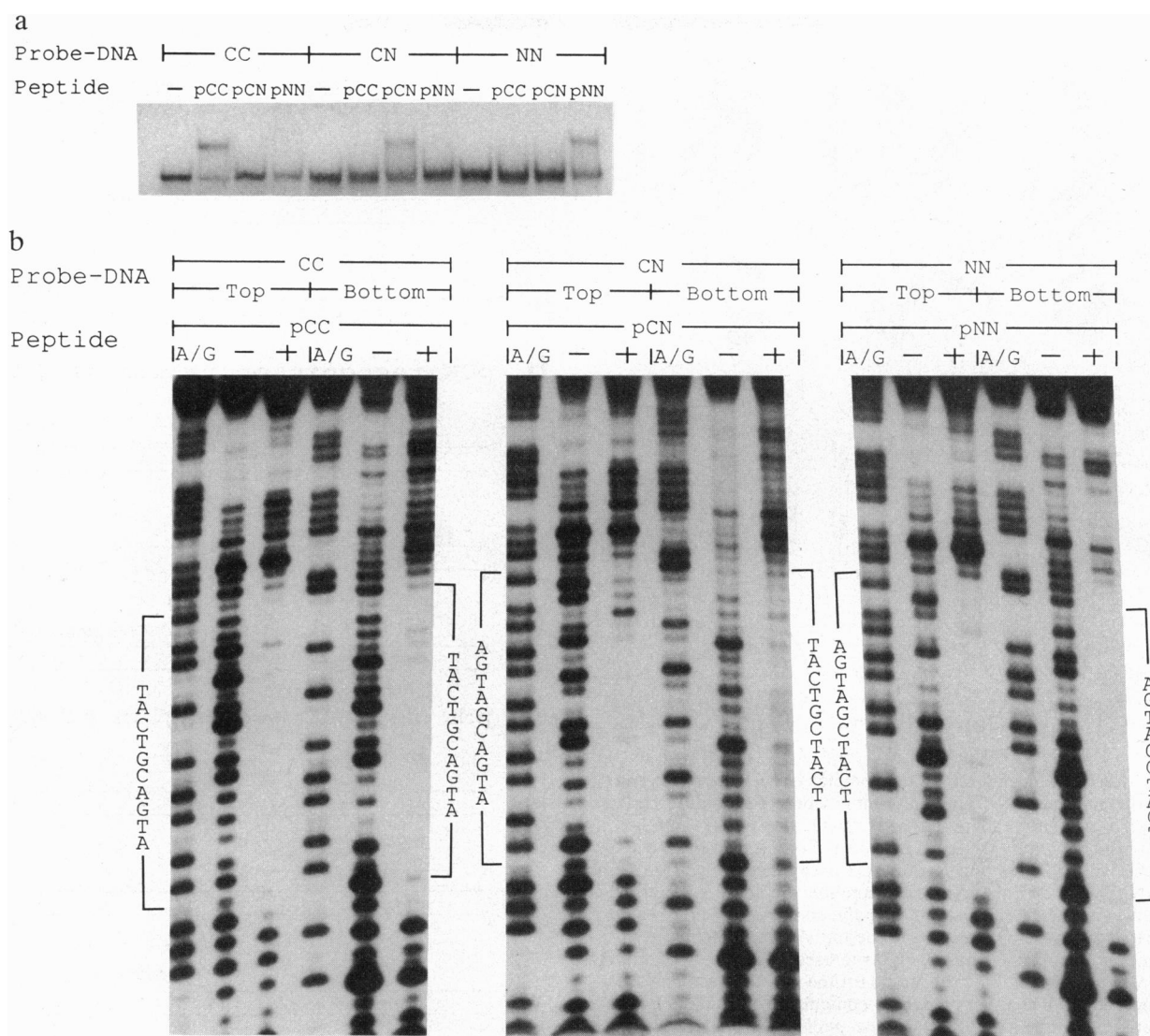
results support the interpretation that the v-Jun homodimer bound to its specific site has bent  $\alpha$ -helices.

## Peptide Design

Using protein stitchery, we have made three kinds of v-Jun (16, 17) homodimers (denoted pCC, pNN, and pCN) and show here that each selectively recognizes the appropriately reorganized DNA binding sites ATGAcgTCAT, TCATcgATGA, and ATGAcgATGA (see Fig. 1). The concept of protein stitchery (15) is that the individual basic arms (half

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

§To whom reprint requests should be addressed at \*.



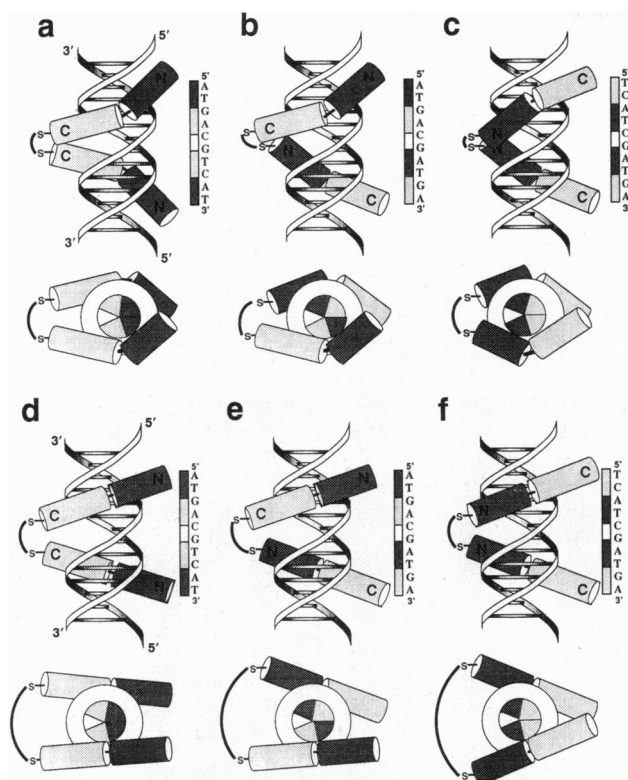
**FIG. 2.** (a) Gel-retardation assay for binding of pCC, pCN, and pNN to the CC, CN, and NN probe DNAs. Binding solution contains bovine serum albumin at 50  $\mu$ g/ml, 10% (vol/vol) glycerol, 20 mM Tris-HCl (pH 7.5), 4 mM KCl, 4 mM MgCl<sub>2</sub>, and the appropriate peptide at 3 nM in a 10- $\mu$ l reaction volume. After 5000 cpm of each 5' <sup>32</sup>P-labeled probe DNA was added, the solutions were stored at 4°C for 40 min and loaded directly on an 8% nondenaturing polyacrylamide gel in Tris/EDTA buffer at 4°C. As determined by titration of the gel shift,  $K_d \approx 2$  nM for pCC/CC,  $K_d \approx 6$  nM for pCN/CN, and  $K_d \approx 4$  nM for pNN/NN, all at 4°C. These results show that each peptide binds specifically to its proposed binding site and not to the other sites. (b) DNase I footprinting assay of pCC, pCN, and pNN peptide with DNA containing the predicted binding sites for pCC, pCN, and pNN, respectively. Footprinting assay solution (in 50  $\mu$ l) contains bovine serum albumin at 50  $\mu$ g/ml, 5% (vol/vol) glycerol, 20 mM Tris-HCl (pH 7.5), 4 mM KCl, 2 mM MgCl<sub>2</sub>, 1 mM CaCl<sub>2</sub>, 20,000 cpm of each 5' <sup>32</sup>P-labeled probe DNA, and the appropriate peptide at 50 nM. This solution was stored at 4°C for 40 min. After 5  $\mu$ l of DNase I diluted in 1 $\times$  footprinting assay buffer was added, the solutions were stored 1 min more at 4°C. DNase I digestion was stopped by addition of 100  $\mu$ l of DNase I stop solution containing 15 mM EDTA (pH 8.0), 100 mM NaCl, and sonicated salmon sperm DNA at 40  $\mu$ g/ml. This mixture was phenol/chloroform-extracted, ethanol-precipitated, and washed with 70% (vol/vol) ethanol. The pellet was resuspended in 5  $\mu$ l of formamide loading buffer, denatured at 90°C for 4 min, and analyzed on 10% polyacrylamide sequencing gel (50% urea). These results show that each peptide specifically binds to the proposed binding site and protects the whole site except for the case of pCN/CN, which shows some incomplete protection on the binding site. This exception is explained as due to binding to semispecific (half) sites by single arms as discussed in the text (see Fig. 4).

sites) of the dimer and the individual half sites of the DNA can be recombined or stitched together in various sequences to form new proteins selective for binding to the new DNA sites. Thus, we use here the recognition helix v-Jun-br of Fig. 1a with a cysteine linker at either the N (v-Jun-N) or the C (v-Jun-C) terminus. These can be combined to form either pNN, pCC, or pCN dimers as illustrated in Fig. 1c. Formation of pNN and pCC (via pathway I) is straightforward since each involves dimerization of identical monomers. To ensure formation of pCN, the cysteine at the C terminus of v-Jun-C was reacted with excess 2,2'-dithiodipyridine to form thiopyridyl-(v-Jun-C) (18, 19) and then coupled with the cysteine at the N terminus of v-Jun-N to form the pCN dimer (v-Jun-

C)-S-S-(v-Jun-N) (pathway III; Fig. 1c). We also verified pathway II for forming pCC.

## Results and Discussion

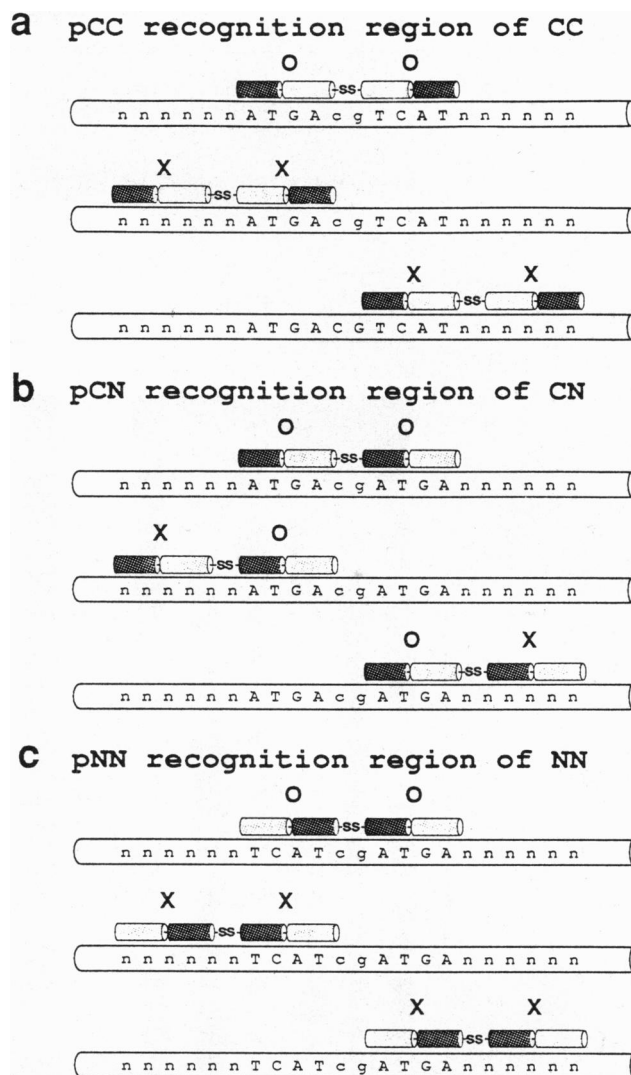
We carried out gel-retardation assays (15) for each of the three peptide dimers with oligonucleotides (Fig. 1b) corresponding to each of the three proposed binding sites. These results (Fig. 2a) show that each dimer recognizes the appropriate binding site specifically with no detectable binding to the other sites. It is important to note that this strong preference for dimer occurs even though all oligonucleotides contain proper sites for binding a single arm of each dimer.



**FIG. 3.** Schematic diagram for the complex between peptides and their corresponding DNA sites assuming a bent recognition helix (a–c) and a straight recognition helix (d–f). (a and d) Complex between pCC and probe DNA CC. (b and e) Complex between pCN and probe DNA CN. (c and f) Complex between pNN and probe DNA NN. The linker connecting two monomers indicates a disulfide bond between the cysteines attached to the end of peptides. In each case, the side view is on the top and the top view is on the bottom. Outer and inner circles of the top view represent the outer and inner major groove surfaces of the top strand for the proposed binding site projected onto an imaginary plane perpendicular to the axis of DNA and running through the center of the peptide and binding site. Shading is used with the peptide and DNA contacts to ease the tracking of these regions in different cases. This diagram shows that a bent recognition helix can contact the same 4 bases for all three peptide dimers, while a linear recognition helix would contact different bases in the three peptide dimers. (This diagram is not meant to imply an exact correlation between where the basic region is bent and where the bases are positioned.)

Therefore, at 3 nM peptide concentration the dimer does not make a stable complex with DNA unless both arms in the dimer recognize their proper sites. This implies cooperation between the monomers in recognizing the binding site (20). Since all three dimers have similar (2–6 nM) binding affinities with their own sites and since all three lead to the same length region protected from DNase I digestion (see below), we conclude that (i) all three cases involve similar conformations in the complex between DNA and peptide, and (ii) the monomer arm retains the same contact region in various dimers; this occurs despite the changing orientation of the monomers in the various peptide dimers (15).

There are two major models for the bound conformation of leucine zipper protein to the specific site. One is the induced helical fork model (13), which proposes a straight single  $\alpha$ -helix for the basic region, and the other is a scissors grip model (5) which proposes a bent  $\alpha$ -helix for the basic region. The recent x-ray crystal structure (21) for the complex of GCN4 containing only the basic and leucine zipper region and DNA-containing GRE site showed that the basic region of each protein has a straight  $\alpha$ -helix conformation recognizing



**FIG. 4.** Specific binding of protein at and near the corresponding DNA binding site. (a) Complex pCC/CC. (b) Complex pCN/CN. (c) Complex pNN/NN. O represents specific binding; X represents nonspecific binding. pCN/CN has one specific binding site and two sites for semispecific (half site) binding near its (nonpalindromic) binding site. However, pCC/CC and pNN/NN do not allow semispecific binding near their (palindromic) binding sites.

each half site of the dimer binding site. There was no DNA bending caused by protein binding (21). However, there remain many problems with assuming that the basic region is in all cases a straight  $\alpha$ -helix: (i) The bases flanking the active site affect the binding of leucine zipper protein even though the crystal structure shows no direct contacts with protein (21). (ii) Gel electrophoresis experiments using Jun homodimer and Jun-Fos heterodimer showed that Jun and Fos induce DNA bending in opposite directions upon binding to their site (22), whereas GCN4 does not induce DNA bending (21, 27). (iii) Even though GCN4-br (a peptide containing the basic region of GCN4 protein) showed no specific binding (for details see ref. 14), we find that the monomer v-Jun-br (a peptide containing only the basic region of v-Jun; see Fig. 1a) specifically binds to the dimer site and shows the same protection as the dimer. Our conclusion then is that there is no universal model for the DNA-bound conformation of the basic region of leucine zipper proteins. Whether it is linear (as in GCN4) or bent (as in Jun) depends on the specific primary sequence and the properties of the solutions (stabilizers, pH, etc.) used in the experiments.

The result that all three dimers (pCC, pNN, and pCN) bind strongly to the appropriate combination of oligonucleotide sites implies that the helical binding arm is bent (5, 22) (see Fig. 3). Our argument is as follows. The optimum binding site for both Jun homodimer and the Jun-Fos heterodimer is known to be ATGAcTCAT or ATGAcgTCAT, where the inner 7 or 8 bases play an important role in recognition (23, 24). The x-ray crystal structure of GCN4 bound to DNA leads to straight  $\alpha$ -helices, which have direct contacts with only the inner 7 bases of the GRE site (ATGACTCAT). Thus, each arm recognizes the half-site (gATGAc or gTCATc) of the dimer binding site asymmetrically. If the same were true for v-Jun and if the same contacts are maintained between the protein and bases for the bound conformations of pCC/CC, pNN/NN, and pCN/CN (as expected since the binding constants and protection are the same), then the orientations of the binding arms would have very different orientations (Fig. 3 *e* and *f*). This should result in different protection from DNase I digestion (not observed). In addition, for the pNN/NN complex, this would lead to N termini of the two arms too distant to be connected by the added linker, GGCCGG. The alternative to Fig. 3 *d-f* is for each dimer to have the same angle (as in Fig. 3*d*). Thus, the actual contact region would not be equivalent in the three cases and it would be difficult to explain the gel retardation and footprinting results. Thus, we conclude that for v-Jun the basic region becomes bent upon binding to the DNA.

On the other hand, with the recognition helix bent roughly at the middle of the helix (as indicated in Fig. 3 *a-c*), it is plausible that the contact regions are ATGAcgTCAT for pCC, TCATcgATGA for pNN, and ATGAcgATGA for pCN. This leads to equivalent contact regions in all three cases and to the roughly equivalent binding energies apparent in Fig. 2*a*. In addition, footprinting (15) of the three peptide dimers (Fig. 2*b*), each with the appropriate oligonucleotide dimer, suggests that the complexed peptide dimers protect the full specific site (all 10 bp) from DNase I digestion. These results strongly support the bent recognition helix model for the basic dimers considered here and hence also for the leucine zipper parent proteins (21, 22).

For the pCN/CN complex, footprinting (Fig. 2*b*) shows incomplete protection on the binding site and partial protection on the bases flanking the binding site, whereas for pCC/CC and pNN/NN this does not happen. This occurs even though gel-retardation assays indicate specific binding for all complexes. Our explanation of this (Fig. 4) suggests why palindromic sequences are so common for selective binding of regulatory proteins (25, 26). This reasoning is supported by recent results we have observed showing that (i) the monomer of v-Jun containing only the basic region (v-Jun-br) specifically protects both pCC and pNN binding sites identically to the protection provided by the dimers pCC and pNN, respectively; (ii) at 3 nM concentration, gel retardation showed that pCC (and pCN) has lower binding affinity for the DNA probe carrying a sequence of cgATGAcgTCATcgTCATcg (containing pCC and pCN binding sites overlapping half of each dimer binding site in the center) than for CC (and CN) probe DNA. These results imply that the half site, gTCATc (or gATGAc), added next to the pCC (or pCN) binding site interferes with the binding of pCC (or pCN) to the dimer binding site (because the half site can be used as a binding site for each arm of the dimer if the orientation between the site and arm fits). Details of these results will be published elsewhere. Fig. 4 indicates the strength of binding for all three peptide dimers at or near their DNA recognition sites. Here, O represents good binding, while X represents nonspecific binding. The palindromic sites for pNN and pCC lead to binding only when the protein is exactly at the recognition site, whereas pCN can recognize both full site (both arms bound) and half sites (one arm bound). In gel

retardation and DNase I footprinting, semispecific binding competes with specific binding. This occurs because one arm of the semispecifically bound peptide would cover half of the specific binding site, preventing another dimer from binding and providing full protection. This explains (i) why gel-retardation assays (Fig. 2*a*) show lower binding affinity for the pCN/CN complex compared to the pCC/CC and pNN/NN complexes and (ii) why footprinting assays (Fig. 2*b*) show incomplete protection on the binding site and partial protection on a few bases flanking the binding site. Such semispecific binding interferes with the site-specific binding and would eventually result in low production and abnormally slow growth. However, gel retardation shows no detectable nonspecific or semispecific binding at low peptide concentration, indicating that semispecific binding is significantly weaker than specific binding. After dimerization, the proteins suitable for palindromic dimer binding sites avoid semispecific DNA binding, leading to more selective recognition of the specific sites. Thus, palindromic dimer binding sites provide a good design for selective molecular recognition and for further flexibility the link can align sites (Fig. 3) to modify recognition.

The results on the three dimers considered here provide encouragement that this protein stitchery approach is feasible for designing and synthesizing proteins to recognize long DNA sequences. Thus, for trimers to recognize 15-bp sequences, we are using an approach similar to that of Fig. 1*c* involving appropriate use of cysteine linkages and transfer activators. It seems possible to design proteins for 20 bp and longer.

In summary, we find the following: (i) Protein stitchery of v-Jun leads to three dimers (pCC, pNN, and pCN), each of which binds specifically to the appropriate rearrangement of DNA sites. Thus, there is cooperation between the two monomers of the dimer in binding to DNA, which depends on the relative orientation of two monomers in the dimer. (ii) These results provide strong support for the bent  $\alpha$ -helix model of the basic region when bound to DNA. (iii) These results provide an explanation for the advantage of dimerization and the use of palindromic sites in the site-selective binding of proteins to DNA. (iv) These results show protein stitchery to be useful for establishing the conformation and mechanism for binding of proteins to their DNA binding sites.

This is contribution no. 8776 from the Division of Chemistry and Chemical Engineering, California Institute of Technology. We thank Prof. Peter Kim for helpful criticisms. This work was supported by a grant from the Department of Energy-Advanced Industrial Concepts Division. The facilities of the Materials and Molecular Simulation Center are also supported by grants from the National Science Foundation-CHE 91-100289, the National Science Foundation-Advanced Computing, the Department of Energy-Chemical Biology, Allied Chemical, Asahi Chemical, Asahi Glass, Chevron, BF Goodrich, BP America, and Xerox.

- Curran, T. & Franza, B. R., Jr. (1988) *Cell* **55**, 395-397.
- Kouzarides, T. & Ziff, E. (1989) *Nature (London)* **340**, 568-571.
- Turner, R. & Tjian, R. (1989) *Science* **243**, 1688-1694.
- Gentz, R., Rauscher, F. J., Abate, C. & Curran, T. (1989) *Science* **243**, 1695-1699.
- Vinson, C. R., Sigler, P. B. & McKnight, S. L. (1989) *Science* **246**, 911-916.
- Neuberg, M., Schermann, M., Hunter, J. B. & Muller, R. (1989) *Nature (London)* **338**, 589-590.
- Ransone, L. J., Visvader, J., Wamsley, P. & Verma, I. M. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 3806-3810.
- Ransone, L. J., Wamsley, P., Mosley, K. L. & Verma, I. M. (1990) *Mol. Cell. Biol.* **10**, 4565-4573.
- Agre, P., Johnson, P. K. & McKnight, S. L. (1989) *Science* **246**, 922-926.
- Neuberg, M., Adamkiewicz, J., Hunter, J. P. & Muller, R. (1989) *Nature (London)* **341**, 243-245.
- Weiss, M. A. (1990) *Biochemistry* **29**, 8020-8024.

12. Patel, L., Abate, C. & Curran, T. (1990) *Nature (London)* **347**, 572–575.
13. O'Neil, K. T., Hoess, R. H. & DeGrado, W. F. (1990) *Science* **249**, 774–778.
14. Talanian, R. V., McKnight, C. J. & Kim, P. S. (1990) *Science* **249**, 769–771.
15. Park, C., Campbell, J. L. & Goddard, W. A., III (1992) *Proc. Natl. Acad. Sci. USA* **89**, 9094–9096.
16. Kaki, Y., Bos, T. J., Davis, C., Starbuck, M. & Vogt, P. K. (1987) *Proc. Natl. Acad. Sci. USA* **84**, 2848–2852.
17. Bos, T. J., Bohmann, D., Tsuchie, H., Tjian, R. & Vogt, P. K. (1988) *Cell* **52**, 705–712.
18. Corey, D. & Schultz, P. G. (1987) *Science* **238**, 1401–1403.
19. Zuckermann, R., Corey, D. & Schultz, P. G. (1987) *Nucleic Acids Res.* **15**, 5305–5321.
20. Abate, C., Luk, D., Gagne, E., Roeder, R. G. & Curran, T. (1990) *Mol. Cell. Biol.* **10**, 5532–5535.
21. Ellenberg, T. E., Brandl, C. J., Struhl, K. & Harrison, S. C. (1992) *Cell* **71**, 1223–1237.
22. Kerpolla, T. K. & Curran, T. (1991) *Science* **254**, 1210–1214.
23. Ryseck, R. & Bravo, R. (1991) *Oncogene* **6**, 533–542.
24. Schule, R., Umesono, K., Mangelsdorf, D. J., Bolado, J., Pike, J. W. & Evans, R. M. (1990) *Cell* **61**, 497–504.
25. Pabo, C. O. & Sauer, R. T. (1984) *Annu. Rev. Biochem.* **53**, 293–321.
26. Landschulz, W. H., Johnson, P. F. & McKnight, S. L. (1988) *Science* **240**, 1759–1764.
27. Gartenberg, M. R., Ampe, C., Steitz, T. A. & Crothers, D. M. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 6034–6038.